Introduction to Predictive Maintenance Application using Machine Learning: The Case of the Injection System of a Diesel Engine

Zineb Znaidi
ENSET Mohammedia, University Hassan II, Morocco
zineb.znaidi@gmail.com (corresponding author)

Moulay El Houssine Ech-Chhibat
ENSET Mohammedia, University Hassan II, Morocco
echchhibate@gmail.com

Azeddine Khiat
ENSET Mohammedia, University Hassan II, Morocco
khiat@enset-media.ac.ma

Mounir El Khiate
ENSET Mohammedia, University Hassan II, Morocco
mounir.elkhiate@gmail.com

Hassan Samri
ENSET Mohammedia, University Hassan II, Morocco
samri@enset-media.ac.ma

Laila Ait El Maalem
ENSET Mohammedia, University Hassan II, Morocco
l.aitelmaalem@gmail.com

Received: 12 October 2024 | Revised: 20 November 2024 and 11 December 2024 | Accepted: 12 December 2024

 DOI: https://doi.org/10.48084/etasr.9250

ABSTRACT

Diesel engines are crucially important in various fields, particularly in the automotive sector, as they ensure a reliable supply of mechanical energy. However, injection system failures, which are among the most recurrent failures, can lead to performance deterioration and increased pollutant emissions and maintenance costs. Therefore, adopting an effective maintenance strategy to analyze and predict such failures would significantly improve the efficiency of these engines. Based on collected data from engines by reliable sensors, the application of predictive maintenance coupled with a machine learning model allows effective prediction of failures for optimal appropriate maintenance. This study presents an approach to diagnosing the injection system of automotive diesel engines using a test bench based on data from temperature sensors installed on engine cylinders. These temperature data exhibit unusual variations in the event of an injection system failure. The Random Forest (RF) algorithm was employed to analyze these data and establish a clear relationship between cylinder temperatures and failure. The proposed model can detect failures associated with the injection system. Performance evaluation, particularly after parameter tuning, underscores the model's efficacy, achieving an accuracy exceeding 97%.

Keywords-diesel engine; injection system failure; predictive maintenance; machine learning; random forest

I. INTRODUCTION

Despite advances in renewable energy and electric vehicles, diesel engines continue to be a technology of paramount importance in the automotive industry. This is primarily due to their advantages and flexibility of use, including their energy efficiency, highly optimal power-to-weight ratio, and ability to operate with various fuel types. The challenging operating conditions and the severity of the environments in which these engines operate pose significant demands for high power and robust components [1]. This makes them more prone to failures and mechanical breakdowns in various engine organs and circuits [2]. In particular, injection system failures are becoming increasingly common, requiring the adoption of an effective maintenance program and the implementation of an efficient diagnostic procedure [3]. This is essential for preventing unplanned shutdown times resulting from such defects, prolonging their lifetime and improving their reliability.

Numerous maintenance-related research has been carried out to identify an effective strategy to enhance the lifetime of diesel engines and to make an optimal choice of the diagnostic process. In this context, predictive maintenance has emerged as one of the most promising topics, particularly compared to traditional preventive methods [4], especially in addressing diesel injection failures. The effectiveness of the maintenance approach relies on the use of new-generation approaches, particularly those that focus on machine learning and artificial intelligence models [5]. These techniques and models enable the prediction of failures in injection systems and other components and circuits in diesel engines, relying on state monitoring techniques, prognostic modules, decision-making, and fault diagnosis [6]. This allows the implementation of a real-time adjustable maintenance plan, intervening in the faulty element as closely as possible to its malfunctioning period. Tracking the degradation of the system, this approach reduces the frequency of breakdowns while optimizing the frequency of preventive actions. Regarding the above information, a study was conducted to implement predictive maintenance in a transportation fleet maintenance workshop to transform current corrective/preventive maintenance into predictive.

II. STATE-OF-THE-ART APPROACHES

Following the development of instrumentation and automation technology, a variety of methods, including analysis of thermodynamic parameters [7], vibrational analysis [8, 9], oil analysis [10], as well as tools and machine learning models for prognosis and monitoring, have recently been addressed in the literature to identify faults in diesel engines, particularly those related to the injection system. In [11], the aim was to identify anomalies in the cylinder injectors in a small diesel engine as soon as an incorrect amount of fuel is injected into one or more cylinders. This was achieved by employing machine learning approaches and striving to identify the most relevant measures for detection, comparing various techniques to select the most suitable one. Optimal performance was observed when anomalies were detected using measurements from both the Electronic Control Unit (ECU) and additional sensors installed on the dedicated test bench. In this configuration, the accuracy of the Ensemble Discriminant Analysis (EDA) algorithm reached 95%.

In [12], Random Forest (RF) regression models and artificial neural networks with Multilayer Perceptron (MLP) were employed to predict failures in diesel engines. In this context, a quantitative severity analysis was implemented to assess the health of diesel engines during the deterioration process by calculating the absolute value of the severity of the failure. This was achieved by combining signal processing techniques and machine learning-based regression. In [13], vibrational analysis was used to predict diesel engine failures on board ships of the Portuguese Navy. This involved conducting a study based on the collection of actual measurements of mechanical vibrations taken on these engines under operational conditions. This study also proposed a graphical approach focused on bi-planes for vibration monitoring in the context of Condition-Based Maintenance (CBM), allowing for the simultaneous display of vibration frequencies and their measurement points.

In [14], a dimensionally adjusted thermodynamic model was adopted to simulate the failure of a four-stroke marine diesel engine. With high precision and reliability, this model generated a significant range of typical thermodynamic faults and failures in diesel engines using RF and MLP neural networks. In [15], an approach was presented to establish an individual predictive maintenance system for diesel engines in railway vehicles, focusing in particular on the maintenance of the injection system and turbochargers. A machine learning technique was employed to determine the probable moment of failure. As a result, the main failure categories specific to a type of engine were identified. An in-depth analysis reveals that the lifespan of the turbochargers and the injection system are not fully utilized.

III. IPRIORITIZATION OF DIESEL ENGINE FAILURES

To prevent breakdowns, optimize operational availability, reduce costs, and improve the overall performance of diesel engines, the implementation of predictive maintenance has recently become the trending action among industrialists to replace conventional methods (preventive and corrective). However, this maintenance approach requires significant investment costs, complex establishment processes, and reliable data. Therefore, predictive maintenance for diesel engines should focus primarily on identifying the most critical components to build an appropriate prognostic model. This study focused first on the diagnostic results of 96 similar diesel engines, intended for semitrailer trucks, with six cylinders arranged in a V-shape, brought in for repair during the year 2023. This study was based on the maintenance intervention history extracted from the various maintenance records specific to each of the engines to identify the different failures encountered and the corrective operations carried out. The analysis of these documents revealed the different failure types of these engines, as well as their occurrence rates. Figure 1 shows the results of the breakdown analysis of the different engine components.

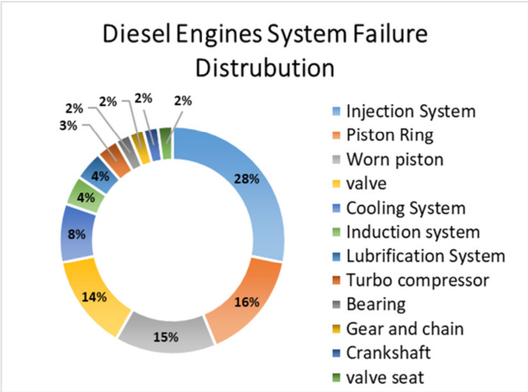

Fig. 1. Diesel engine component failure distribution.

Examining the distribution in Figure 1 reveals elements that are important for the functioning of the engines under study. In particular, 80% of failures were caused by injection systems, piston rings, worn pistons, valves, and cooling systems. These five elements must be of extreme significance in developing a plan of intervention for predictive maintenance and spare parts management. Additionally, compared to the other four key items, the injection system failure alone accounted for more than 36% of the failure rates. Therefore, there is a need to prioritize the injection system, as it significantly affects the functionality, dependability, and cost of repairs. Furthermore, the final findings of this analysis are in agreement with the literature. In [16, 17], a classification of the most frequently occurring diesel engine failure modes showed that the injection system accounted for 27% of failures in diesel engines, as shown in Figure 2.

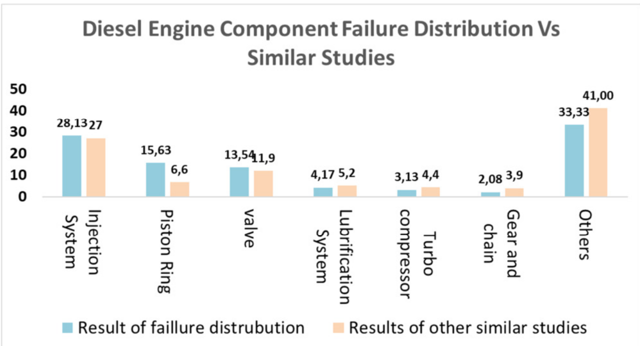

Fig. 2. Diesel engine component failure distribution vs similar studies.

IV. EXPERIMENTAL STUDY AND MACHINE LEARNING MODEL CONSTRUCTION

A. *Objective of the Study*

For a complete diagnostic cycle on the test bench, each engine needs more than 6 hours (fuel supply, ignition, cooling, injection, etc.). Since injection system failures are the most frequent, this study initiated the application of predictive maintenance for failure detection based on maintenance intervention by extracting every possible pattern between the failure and the relative temperature data collected from the sensors installed in the injection system.

B. *Description of the Studied System*

The test bench used for engine diagnosis comprises three main functions:

- Monitoring function: Based on a Human-Machine Interface (HMI) facilitating the use of the test bench, it allows the user to visualize in real-time the evolution of the different variables measured [18] (speed, torque, pressures, temperature, power, etc.) as well as the associated commands.
- Command and control function: Generates sequenced control signals for the engine (injection, ignition, etc.) and ensures the acquisition of data from the sensors.
- Sequencing function: An electronic stage that performs the sequencing of the engine control using an optical encoder and measurements linked to the crankshaft of the heat engine.

The tested engine is a four-stroke diesel engine tractor-trailer (truck) with six V-cylinders, a turbocharger, and a common rail injection system that uses a high-pressure fuel pump to pressurize fuel and deliver it to a shared reservoir the common rail. This ensures a consistent fuel supply to each injector, which precisely delivers fuel into the cylinders. The ECU manages injection timing, fuel quantity, and pressure in real time, optimizing engine performance and efficiency. Table I shows the technical specifications for this engine.

TABLE I. DIESEL ENGINE'S TECHNICAL SPECIFICATIONS

Engine Parameters	Value
Horsepower	620 Hp
Cylinders	6
Valve control	On the head cylinder
Cylinders valves	2 valves
Diameter of the cylinder	5.750 in
Piston stroke	5.750 in
Total displacement	149.1 cu in.per cylinder
Compression ratio	16:1
Ignition order	1–5–3–6–2–4
Governed speed	2400 rpm
Cooling water temperature	80 to 100 °C

The temperature sensors are K-type thermocouples as seen in Figure 3. The technical characteristics and the accuracy of the sensors are mentioned in Table II. Figure 4 shows an example of the installation position of the temperature sensors on the cylinder. Similar ones are installed in each cylinder.

Fig. 3. Temperature sensor (thermocouple type K).

TABLE II. TEMPERATURE SENSOR CHARACTERISTICS

Characteristics	Indication/value
Metal A (+)	Chrome
Metal B (-)	Alumel
Operating temperature range (continuous use)	0 to +1250°C
Operating temperature range (intermittent use	-180 to +1350°C
Coef. Seebeck α(µV/°C)	39.45 µV/°C to 0°C
Standard error	0.75% to 2.2%

Fig. 4. Temperature sensors installation position.

C. *Experiments and Engine Temperature Analysis*

The manufacturer of these engines provided parameters for the test investigation of injection system defects, which was conducted under conditions of changing speed and torque. Figure 5 displays the applied speeds and torque values for all engines.

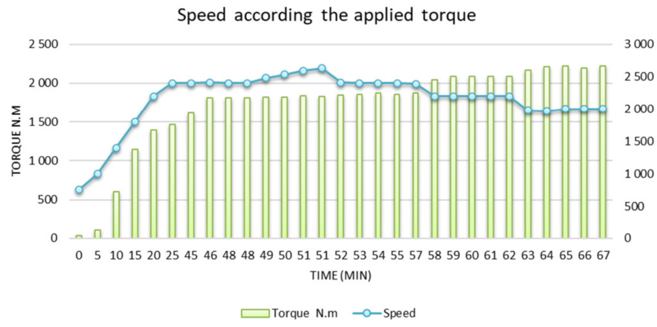

Fig. 5. Applied speed and torque values during each test.

After each sensor installation in the various positions on the top of the cylinders and before starting the thermal engine test phase, the self-test operation of the sensors is carried out to ensure their functionality to reduce temperature measurement errors [19, 20]. The experimentation began with the establishment of the reference curve. In this case, it was assumed that the curve of the new engines would serve as a reference since these new engines would have undergone thorough testing and validation at the manufacturer to ensure their reliable operation. Three new engines were studied to establish the reference curve, using the Mustache box shown in Figure 6, to provide a comprehensive view of the data distribution and aid in understanding the central tendency and variability of the dataset. Table III shows statistical data for the three new engines, which are mostly similar with an acceptable range for each statistical characteristic.

TABLE III. NEW ENGINES' STATISTICAL DATA

	New_Engine1	New_Engine2	New_Engine3	Range
Average	904	931	905	26
Median	1056	1079	1055	23
Min	186	187	185	2
Max	1140	1167	1140	27

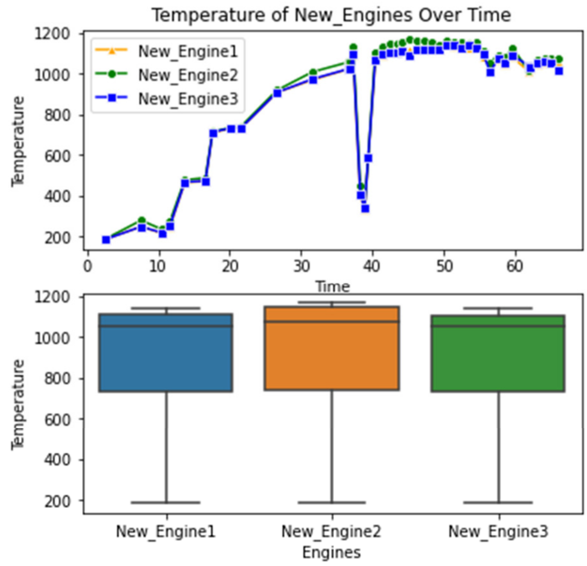

Fig. 6. Reference curve with related Mustache box.

Subsequently, the experiment began with the diagnosis on the test bench of engines without any failure, from the six sensors installed on the cylinders. Temperature data were recorded in an 80-minute test. Figure 7(a-c) shows examples of temperature evolution over time. Then, the diagnostics were conducted on the test bench for other engines with injection system failures, and Figure 7(c-f) shows examples of temperature evolution over time.

The comparison of each engine curve with the reference shows that engines without related failure to the injection system are still generally similar to the reference curve. However, engines with injection system failure present a different temperature distribution, namely after the typical temperature reduction related to the applied torque. This temperature pattern change constitutes an important key to making out engines with injection system failure during the stage of the machine learning application [21].

D. *Fault Detection based on Random Forest (RF)*

1) *RF Applied to Predictive Maintenance*

As machine learning continues to advance, its applications in industrial settings have become increasingly valuable [22]. One such application is the use of machine learning techniques for predictive maintenance, aiming to forecast equipment failures and optimize maintenance schedules. Among the various available machine learning algorithms, RF has emerged as a powerful tool for addressing predictive maintenance challenges [23]. This study presents a comprehensive introduction to the application of RF for predictive maintenance, highlighting its theoretical foundations and practical implementation using Python along with the key characteristics that make it a suitable choice for this domain. Then, the process of data mining and model training is described, emphasizing the importance of domain-specific knowledge in optimizing the RF model for predictive maintenance tasks. RF employs the concept of decision trees, constructing a collection of decision trees and aggregating their results to generate the final prediction [24]. Every decision tree within an RF is constructed using random subsets of data, and each tree is trained on a portion of the entire dataset [25].

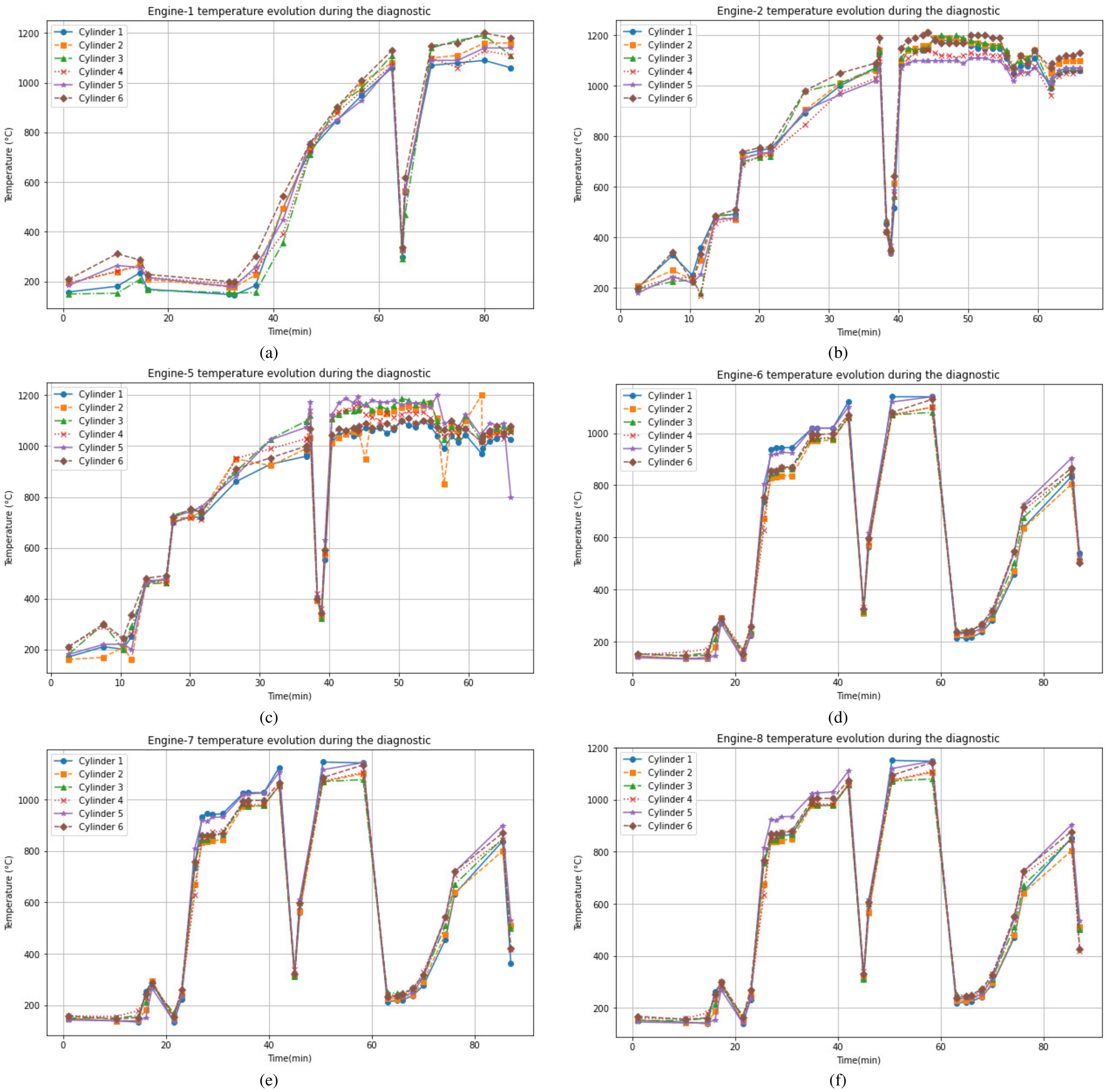

Fig. 7. Temperature evolution in: (a) Engine 1, (b) Engine 2, (c) Engine 3, (d) Engine 4, (e) Engine (5), (f) Engine (6).

The outcomes of all decision trees are amalgamated to derive the ultimate forecast:

$$\hat{y} = \arg max_c \sum_{j=1}^{T} 1(h_j(x) = c) \quad (1)$$

where $\hat{y}$ is the predicted class for the sample x, and $argmax_c$ is the value of c (class) that maximizes the following function:

$$\sum_{j=1}^{T} 1(h_j(x) = c) \quad (2)$$

This is the sum of votes for class c among the T trees and $h_j(x)$ is the prediction of the j tree for the sample x.

The Gini index measures the extent of impurity or disorder in a dataset [26] and is calculated for each node in a decision tree to determine the best split that results in the highest homogeneity of resulting nodes. The formula for the Gini index is given by:

$$Gini(D) = 1 - \sum_{i=c}^{CL} Pi^2 \quad (3)$$

where Pi is the proportion of instances belonging to class c in the dataset D, and CL is the number of classes.

RF is well-suited for predictive maintenance applications due to:

- Its ability to handle complex, nonlinear relationships within the data, as well as its resilience to overfitting [27].
- It can select the most important features from a dataset for improved accuracy. This is done by constructing multiple decision trees and measuring the importance of each feature based on its contribution to accuracy [28].
- Feature selection capability, where many potential sensor inputs and operational variables could be used to predict equipment failures. RF can help identify the critical variables that are most predictive of failures.
- RF is also a flexible and easy-to-use machine learning algorithm that produces great results even without extensive hyperparameter tuning [29].
- The algorithm's ability to handle both classification and regression tasks also makes it suitable for a wide range of predictive maintenance use cases.
- It is known to be robust to overfitting, a common challenge in predictive maintenance where models may be trained on limited historical failure data [30].

2) Data Mining

Data mining contributes to the development of predictive models by identifying relationships and dependencies in data [31]. Machine learning algorithms leverage these insights to make accurate predictions or classifications on new/unseen data. Figure 8 shows the approach adopted for fault detection.

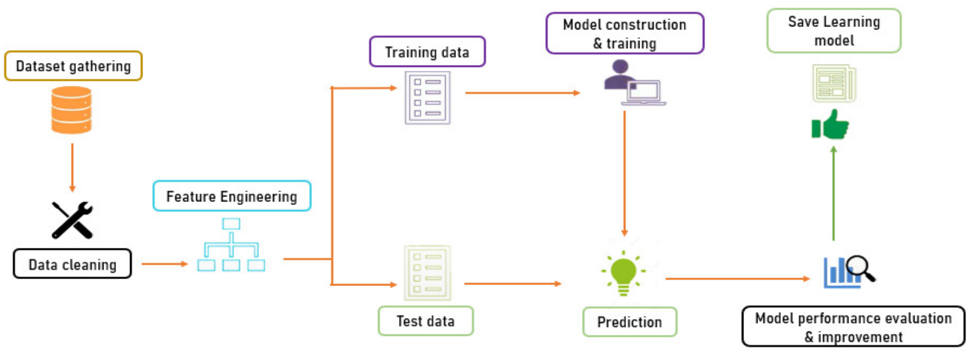

Fig. 8. The proposed method for fault diagnosis in a diesel engine.

The proposed approach to fault diagnosis employs datasets from the bench test for all engines tested. To obtain the most accurate prediction details, the proposed approach consists of five primary stages: data cleaning, selecting optimal features, and training, testing, and evaluating the prediction model. Initially, the datasets were cleaned through the detection of outliers and the filling of missing values. Subsequently, optimal feature selection was carried out to eliminate redundant data and to assess the optimal sensors needed for accurate data collection using a correlation matrix [32]. In addition, a dispersion diagram was used to visualize the relationship between features and the target variable. Then, RF was employed for injection system fault diagnosis for several engines with and without failure. Additionally, evaluation metrics were used to evaluate the accuracy of the model for reliable fault diagnoses in the future.

V. RESULTS AND DISCUSSION

A. Correlation Matrix

A correlation matrix is an effective analytical tool for simultaneously studying associations between multiple variables [33]. In the context of this study, it allows the validation of the sensor's accuracy and defines the optimal number of sensors to be installed in the systems according to the identified patterns and dependencies that will be concluded [34]. This study sought to quantify the strength and direction of these relationships, providing valuable insight into the factors that may influence the sensor's performance.

Figure 9 shows the correlation matrix of the temperature readings from six sensors placed in each cylinder of the diesel engines, revealing a high correlation coefficient of approximately 0.99 between each pair of sensors. This indicates that each sensor is providing consistent and accurate readings, as expected, and suggests that the temperature across all cylinders is almost identical during the test. This uniformity in temperature readings can be attributed to the impact of the injection pump failure, which affects all cylinders similarly. Consequently, the sensors provide highly correlated data, reflecting the overall thermal response of the engine to failure. In addition, the near-perfect correlation coefficients (0.99) suggest a high level of redundancy in the information provided by these sensors. The strong positive correlation indicates that when one sensor registers a certain value, the other sensors tend to record similar values. This redundancy implies that having all sensors might not significantly improve the overall information gain for this specific failure, but future work will aim to simulate failures related to individual cylinders to understand localized effects, differentiating between systemic issues, such as injection pump failures, and specific cylinder-related problems. This will enhance the ability to accurately diagnose and address specific engine issues.

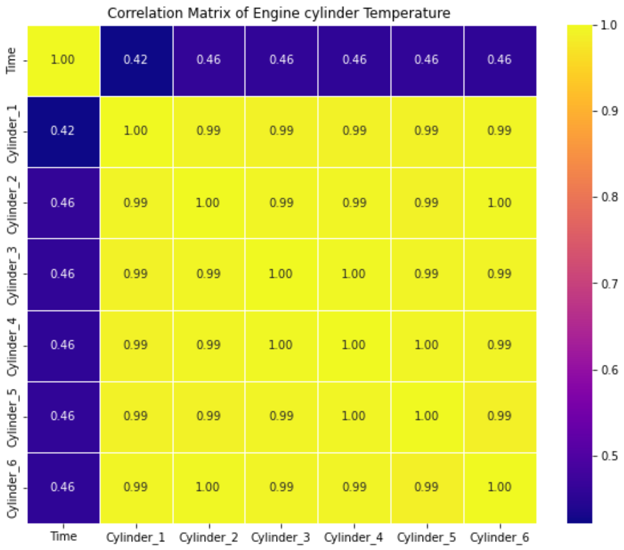

Fig. 9. Correlation matrix.

B. *Dispersion Diagram*

A dispersion diagram, commonly known as a scatter plot, is a graph used to display the relationship between two numerical variables [35]. In a scatter plot, each point represents an observation in the dataset. In this case, Figure 10 presents the temperature evolution in both engines with and without failure. The blue plots show a common pattern for engines with and without failure, and the green plots are related patterns to engines without failure. Finally, the red plot shows specific patterns of the engines with failure. By capturing these patterns, the trees in the RF can better understand the complex structures. This makes the model more robust to variations in the data and improves its ability to generalize to new data.

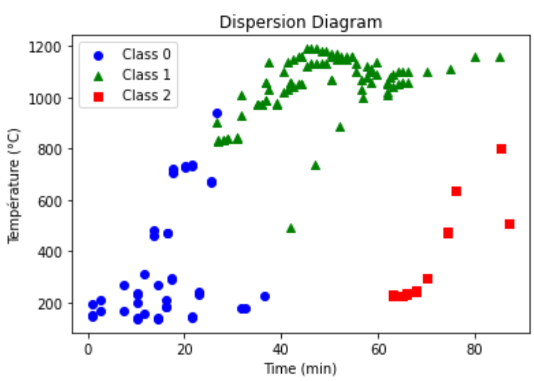

Fig. 10. Dispersion diagram.

C. *Random Forest*

This study aims to predict injection system failures based on sensor data, with three main classes: engine with failure, engine without failure, and common temperature evolution. The bagging concept was applied for sampling. A baseline RF model was trained and evaluated first to establish a reference performance metric. This initial model provided insight into the classifier performance without hyperparameter tuning [36]. Then GridSearchCV, a robust hyperparameter optimization technique [37], was used to exhaustively search through a predefined grid of parameters. The grid included variations in the number of trees ($n\_estimators$), tree depth that refers to the length of the longest path from the root node to a leaf node (max_depth), the minimum number of samples required to split an internal node ($min\_samples\_split$), and the minimum samples required at a leaf node ($min\_samples\_leaf$). GridSearchCV was configured with a three-fold cross-validation. After the grid search was completed, the best combination of hyperparameters was identified based on the highest cross-validated accuracy score. The optimized RF model was then retrained throughout the entire training set using the best parameters and subsequently evaluated on the test set. Figures 11 and 12 show the algorithm flowchart and examples of the obtained trees, respectively.

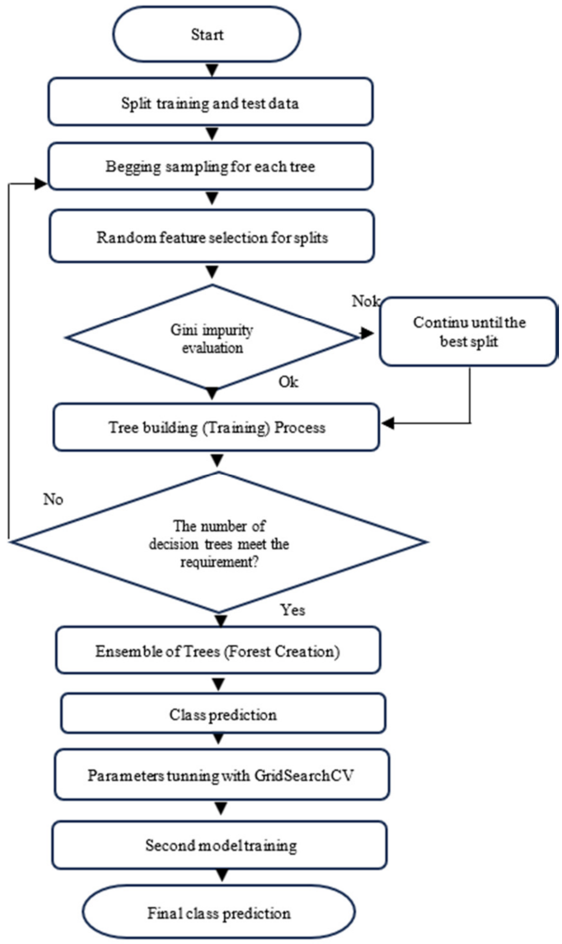

Fig. 11. RF program flowchart.

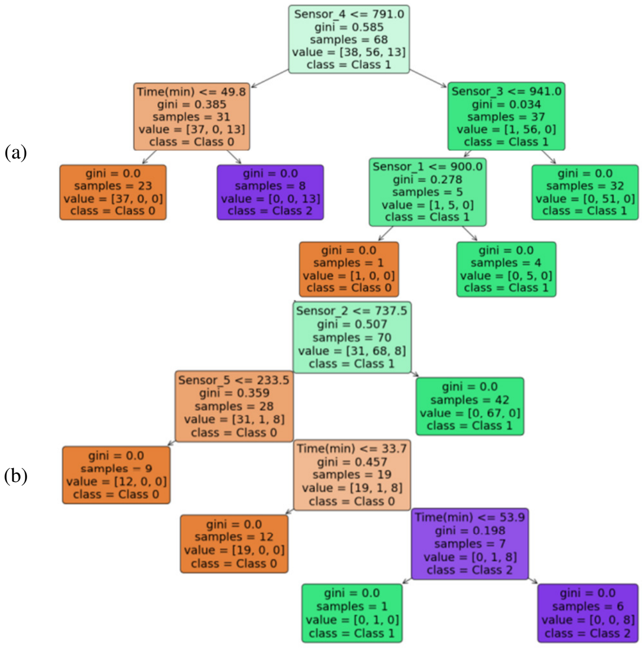

Fig. 12. Examples of obtained trees: (a) Tree 6, (b) Tree 8.

D. Evaluation Metrics

Each hyperparameter is calculated based on the following metrics:

- Accuracy is the proportion of correctly predicted instances out of the total instances [38] and gives a general idea of how well the classifier performs across all classes. TP denotes True Positives, TN denotes True Negatives, FP represents False Negatives, FP denotes False Positives, and TS denotes Total Samples.

$$Accuracy = (TP + TN)/TS \quad (4)$$

- Precision is the ratio of correctly predicted positive observations to the total predicted positives [38].

$$Precision = TP/(TP + FP) \quad (5)$$

- Recall (sensitivity) is the ratio of correctly predicted positive observations to all observations in the actual class [38].

$$Recall = TP/(TP + FN) \ (6)$$

- F1-score is the harmonic mean of Precision and Recall, providing a single metric that balances the trade-off between them [38].

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (7)$$

Evaluation metrics were significantly improved compared to the baseline model. Figure 13 demonstrates the effectiveness of hyperparameter optimization in enhancing the classifier's predictive capabilities. This methodological approach ensures a robust and generalizable model, suitable for deployment in real-world applications.

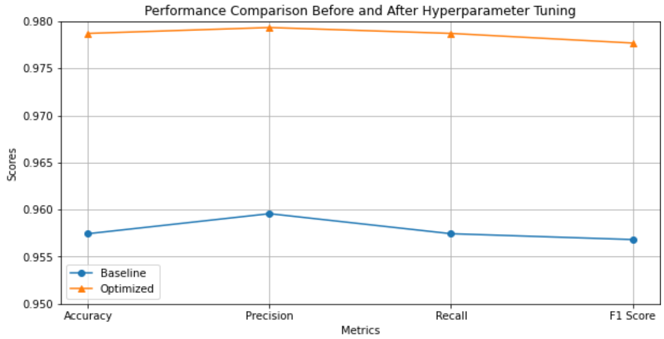

Fig. 13. Model performance before and after hyperparameter tuning.

VI. CONCLUSION AND FUTURE WORK

This study designed and implemented an RF-based machine-learning model to detect failures in the injection system of a diesel engine. This approach aims specifically to identify injector faults and contribute to the implementation of a predictive maintenance strategy. The initial results demonstrated promising performance, with accuracy, precision, and F1-score of more than 95%. To further enhance the model's effectiveness, GridSearchCV was applied for hyperparameter optimization, leading to a significant improvement in the model's overall performance and achieving accuracy exceeding 97%. These results highlight not only the inherent efficiency of the RF algorithm for fault detection applications but also the importance of hyperparameter optimization to achieve superior accuracy. This program represents a robust and reliable tool for detecting injection system failures in diesel engines. These findings underscore the usefulness of supervised learning algorithms in industrial settings, where early fault detection plays a critical role in reducing downtime, improving productivity, and lowering maintenance costs. Future research will aim to predict specific failures related to each injector and develop more complex models to address a broader range of critical failures associated with injection and other subsystems. This will involve integrating more comprehensive data from multiple sources and exploring advanced machine learning techniques to better capture the complex interactions within the engine system. Additionally, future studies will evaluate whether RF remains adequate for these more complex scenarios or whether it could be beneficial to combine it with other machine learning models to achieve optimal performance.

REFERENCES

[1] X. Wen, S. Gierth, M. Rieth, J. H. Chen, and C. Hasse, "Large-eddy simulation of a multi-injection flame in a diesel engine environment using an unsteady flamelet/progress variable approach," *Physics of Fluids*, vol. 33, no. 10, Oct. 2021, Art. no. 105107, https://doi.org/10.1063/5.0065351.

[2] W. Xi, Z. Li, Z. Tian, and Z. Duan, "A feature extraction and visualization method for fault detection of marine diesel engines," *Measurement*, vol. 116, pp. 429–437, Feb. 2018, https://doi.org/10.1016/j.measurement.2017.11.035.

[3] H. Ma, J. Zeng, R. Feng, X. Pang, Q. Wang, and B. Wen, "Review on dynamics of cracked gear systems," *Engineering Failure Analysis*, vol. 55, pp. 224–245, Sep. 2015, https://doi.org/10.1016/j.engfailanal.2015.06.004.

[4] Z. Znaidi, M. E. H. Ech-Chhibat, A. Khiat, and L. A. El Maalem, "Predictive maintenance project implementation based on data-driven & data mining," in *2023 3rd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Mohammedia, Morocco, May 2023, pp. 1–5, https://doi.org/10.1109/IRASET57153.2023.10152915.

[5] S.-Y. Chuang, N. Sahoo, H. W. Lin, and Y. H. Chang, "Predictive Maintenance with Sensor Data Analytics on a Raspberry Pi-Based Experimental Platform," *Sensors*, vol. 19, no. 18, Jan. 2019, Art. no. 3884, https://doi.org/10.3390/s19183884.

[6] W. Prommachan, P. Surin, P. Srinoi, and M. Pipathattakul, "Selection Criteria for Evaluating Predictive Maintenance Techniques for Rotating Machinery using the Analytic Hierarchical Process (AHP)," *Engineering, Technology & Applied Science Research*, vol. 14, no. 1, pp. 13058–13065, Feb. 2024, https://doi.org/10.48084/etasr.6816.

[7] R. Varbanets *et al.*, "Enhanced diagnostic techniques for marine diesel engines: achieving imo decarbonization compliance," *Internal Combustion Engines*, no. 2, pp. 26–36, Sep. 2023, https://doi.org/10.20998/0419-8719.2023.2.04.

[8] M. Zhang, V. Sharma, Y. Jia, A. K. Hossain, and Y. Xu, "Data-driven Approach for Condition Assessment of a Diesel Engine Powered with Various Biodiesels," SAE International, Warrendale, PA, SAE Technical Paper 2023-01–0422, Apr. 2023. https://doi.org/10.4271/2023-01-0422.

[9] A. F. Khalil and S. Rostam, "Machine Learning-based Predictive Maintenance for Fault Detection in Rotating Machinery: A Case Study," *Engineering, Technology & Applied Science Research*, vol. 14, no. 2, pp. 13181–13189, Apr. 2024, https://doi.org/10.48084/etasr.6813.

[10] H. Raposo, J. T. Farinha, I. Fonseca, and L. A. Ferreira, "Condition Monitoring with Prediction Based on Diesel Engine Oil Analysis: A Case Study for Urban Buses," *Actuators*, vol. 8, no. 1, Mar. 2019, Art. no. 14, https://doi.org/10.3390/act8010014.

[11] P. Danti, R. Minamino, and G. Vichi, "Wrong Injection Detection in a Small Diesel Engine, a Machine Learning Approach," *PHM Society European Conference*, vol. 7, no. 1, pp. 87–95, Jun. 2022, https://doi.org/10.36001/phme.2022.v7i1.3311.

[12] D. P. Viana *et al.*, "Diesel Engine Fault Prediction Using Artificial Intelligence Regression Methods," *Machines*, vol. 11, no. 5, May 2023, Art. no. 530, https://doi.org/10.3390/machines11050530.

[13] V. Vairinhos, R. Parreira, S. Lampreia, V. Lobo, and P. Galindo, "Vibration Analysis Based on HJ-Biplots," *International Journal of Prognostics and Health Management*, vol. 9, no. 2, 2018, https://doi.org/10.36001/ijphm.2018.v9i2.2739.

[14] J. A. Pagán Rubio, F. Vera-García, J. Hernandez Grau, J. Muñoz Cámara, and D. Albaladejo Hernandez, "Marine diesel engine failure simulator based on thermodynamic model," *Applied Thermal Engineering*, vol. 144, pp. 982–995, Nov. 2018, https://doi.org/10.1016/j.applthermaleng.2018.08.096.

[15] H. Engbers, S. Leohold, and M. Freitag, "Individual Predictive Maintenance Approach for Diesel Engines in Rail Vehicles," in *Dynamics in Logistics*, 2020, pp. 236–244, https://doi.org/10.1007/978-3-030-44783-0_23.

[16] S. Selcuk, "Predictive maintenance, its implementation and latest trends," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 231, no. 9, pp. 1670–1679, Jul. 2017, https://doi.org/10.1177/0954405415601640.

[17] G. Oskam, "Optimizing diesel engine condition monitoring: Research on diagnostic representation techniques based on in-cylinder pressure measurement," M.S. Thesis, Delft University of Technology, Netherlands, 2014.

[18] E. K. Mounir, S. Hassan, S. Hassan, and Z. Zineb, "An Interface Development based on Internet of Things Approach for Smart Predictive Maintenance implementation: Case of Diesel Engine," in *2024 International Conference on Circuit, Systems and Communication (ICCSC)*, Fes, Morocco, Jun. 2024, pp. 1–6, https://doi.org/10.1109/ICCSC62074.2024.10616750.

[19] R. Radonja, V. Pelić, D. Pavić, and N. Tomac, "Cost efficiency of optimizing automatic temperature control parameters in a diesel engine cooling system on a cruising vessel - A case study," *Journal of Applied Engineering Science*, vol. 18, no. 2, pp. 251–256, May 2020, https://doi.org/10.5937/jaes18-25189.

[20] B. Belhassein, "Caractérisation et modélisation de la thermique d'un moteur Diesel et de sa ligne d'échappement," Ph.D. dissertation, Ecole Centrale de Nantes (ECN), 2014.

[21] O. O. Aremu, R. A. Cody, D. Hyland-Wood, and P. R. McAree, "A relative entropy based feature selection framework for asset data in predictive maintenance," *Computers & Industrial Engineering*, vol. 145, Jul. 2020, Art. no. 106536, https://doi.org/10.1016/j.cie.2020.106536.

[22] G. Anandhi and M. Iyapparaja, "Photocatalytic degradation of drugs and dyes using a maching learning approach," *RSC Advances*, vol. 14, no. 13, pp. 9003–9019, 2024, https://doi.org/10.1039/D4RA00711E.

[23] A. Bakdi, N. B. Kristensen, and M. Stakkeland, "Multiple Instance Learning With Random Forest for Event Logs Analysis and Predictive Maintenance in Ship Electric Propulsion System," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7718–7728, Aug. 2022, https://doi.org/10.1109/TII.2022.3144177.

[24] K. Shah, H. Patel, D. Sanghvi, and M. Shah, "A Comparative Analysis of Logistic Regression, Random Forest and KNN Models for the Text Classification," *Augmented Human Research*, vol. 5, no. 1, Mar. 2020, Art. no. 12, https://doi.org/10.1007/s41133-020-00032-0.

[25] H. A. Salman, A. Kalakech, and A. Steiti, "Random Forest Algorithm Overview," *Babylonian Journal of Machine Learning*, vol. 2024, pp. 69–79, Jun. 2024, https://doi.org/10.58496/BJML/2024/007.

[26] M. I. Prasetiyowati, N. U. Maulidevi, and K. Surendro, "The accuracy of Random Forest performance can be improved by conducting a feature selection with a balancing strategy," *PeerJ Computer Science*, vol. 8, Jul. 2022, Art. no. e1041, https://doi.org/10.7717/peerj-cs.1041.

[27] J. M. dos S. Rodrigues Neto and N. P. Bhandary, "Landslide Susceptibility Assessment by Machine Learning and Frequency Ratio Methods Using XRAIN Radar-Acquired Rainfall Data," *Geosciences*, vol. 14, no. 6, Jun. 2024, Art. no. 171, https://doi.org/10.3390/geosciences14060171.

[28] V. Santilli *et al.*, "The Use of Machine Learning for Inferencing the Effectiveness of a Rehabilitation Program for Orthopedic and Neurological Patients," *International Journal of Environmental Research and Public Health*, vol. 20, no. 8, Jan. 2023, Art. no. 5575, https://doi.org/10.3390/ijerph20085575.

[29] F. Jiang *et al.*, "Estimating the Growing Stem Volume of Coniferous Plantations Based on Random Forest Using an Optimized Variable Selection Method," *Sensors*, vol. 20, no. 24, Jan. 2020, Art. no. 7248, https://doi.org/10.3390/s20247248.

[30] W. Zhang, C. Wu, H. Zhong, Y. Li, and L. Wang, "Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization," *Geoscience Frontiers*, vol. 12, no. 1, pp. 469–477, Jan. 2021, https://doi.org/10.1016/j.gsf.2020.03.007.

[31] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Computer Science*, vol. 2, no. 3, Mar. 2021, Art. no. 160, https://doi.org/10.1007/s42979-021-00592-x.

[32] H. Darvishi, D. Ciuonzo, E. R. Eide, and P. S. Rossi, "Sensor-Fault Detection, Isolation and Accommodation for Digital Twins via Modular Data-Driven Architecture," *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4827–4838, Oct. 2021, https://doi.org/10.1109/JSEN.2020.3029459.

[33] Z. Zhou *et al.*, "Performance Evaluation of Oscillating Wall Stirrer in Reverberation Chamber Using Correlation Matrix Method and Modes Within Q-Bandwidth," *IEEE Transactions on Electromagnetic Compatibility*, vol. 62, no. 6, pp. 2669–2678, Sep. 2020, https://doi.org/10.1109/TEMC.2020.2983981.

[34] A. Salmivaara *et al.*, "Wheel rut measurements by forest machine-mounted LiDAR sensors – accuracy and potential for operational applications?," *International Journal of Forest Engineering*, vol. 29, no. 1, pp. 41–52, Jan. 2018, https://doi.org/10.1080/14942119.2018.1419677.

[35] F. Mohseni and M. Mokhtarzade, "The synergistic use of microwave coarse-scale measurements and two adopted high-resolution indices driven from long-term T-V scatter plot for fine-scale soil moisture estimation," *GIScience & Remote Sensing*, vol. 58, no. 3, pp. 455–482, Apr. 2021, https://doi.org/10.1080/15481603.2021.1906056.

[36] M. Schonlau and R. Y. Zou, "The random forest algorithm for statistical learning," *The Stata Journal*, vol. 20, no. 1, pp. 3–29, Mar. 2020, https://doi.org/10.1177/1536867X20909688.

[37] Z. M. Alhakeem, Y. M. Jebur, S. N. Henedy, H. Imran, L. F. A. Bernardo, and H. M. Hussein, "Prediction of Ecofriendly Concrete Compressive Strength Using Gradient Boosting Regression Tree Combined with GridSearchCV Hyperparameter-Optimization Techniques," *Materials*, vol. 15, no. 21, Jan. 2022, Art. no. 7432, https://doi.org/10.3390/ma15217432.

[38] C. Iwendi *et al.*, "COVID-19 Patient Health Prediction Using Boosted Random Forest Algorithm," *Frontiers in Public Health*, vol. 8, Jul. 2020, https://doi.org/10.3389/fpubh.2020.00357.